Exploratory Data Analysis in a Study of Breast Cancer and the Environment

Steven J. Melly (1),[1] Nancy I. Maxwell (1), Yvette T. Joyce (2), Julia G. Brody (1)
(1) Silent Spring Institute, Newton, MA; (2) Applied Geographics, Inc., Boston, MA

Abstract

In the first phase of the Cape Cod Breast Cancer and Environment Study we used a geographic information system (GIS) as the central management and analysis tool in a detailed cancer surveillance effort and community-level study. We mapped patterns of breast cancer incidence in relation to environmental exposures of concern including infiltration of waste water into drinking water and large-scale historical pesticide use. We developed methods to compensate for some limitations in the data including differences in source scales. Part of our work included using the GIS together with a statistical program for exploratory data visualization. We used the statistical program to explore the effects of using different cutpoints to define categories of both exposure and disease. This exploratory analysis brought together data on US Census units with geographic information on point and non-point sources of environmental pollution from a range of data sources. Results from this first phase of research were used to plan a case-control study that began in the fall of 1998.

Keywords: breast cancer, endocrine disrupters, drinking water, waste water

Introduction

With increasing access to health surveillance data from state cancer registries and other sources, communities across the country are learning how disease rates in their area compare with those in other areas. As new statistics are published, high incidence communities want to know why their rates are high and how to bring them down. The Cape Cod Breast Cancer and Environment Study illustrates how geographic information system (GIS) technology can be used both to develop more accurate and detailed descriptions of disease incidence and to explore the causes. The study is being conducted by the Silent Spring Institute, a nonprofit research organization dedicated to studying the links between women's health and the environment. The Institute is funded by the Massachusetts Department of Public Health.

When the Cape Cod Study began in 1994, Massachusetts Cancer Registry data indicated that age-adjusted breast cancer incidence was significantly higher in a majority of Cape Cod towns than in the state as a whole. Alarmed by these statistics, citizen activists, public health officials, and researchers began sifting through possible explanations. Were high breast cancer rates due to characteristics of women who live on the Cape, better mammography screening, or something about the environment?

Information about the population of Cape Cod suggested that it was substantially similar to the rest of the state. In contrast, the environment of the Cape is obviously

[1] Steven J. Melly, Silent Spring Institute, 29 Crafts Street, Newton, MA 02158 USA; (p) 617-332-4288; (f) 617-332-4284; E-mail: melly@silent.shore.net

different. Nearly all of the population relies for its drinking water on groundwater from a sand and gravel aquifer overlain by sandy soils, so drinking water wells are vulnerable to contamination from septic tanks and other land uses. Historically, pesticide use on the Cape may have been greater than elsewhere because of the large number of cranberry bogs, golf courses, and trees susceptible to gypsy moth and other pests. In addition, the Cape hosts two military facilities. All of these factors pointed to the environment as a possible key to understanding breast cancer on the Cape. Because no *single* factor stood out as a likely cause of the elevated breast cancer incidence, however, it was important to design a study that would allow the exploration of many factors. Considering multiple factors also was important given the complexity of the disease and the possible effects of carcinogens, tumor promoters, and genetics.

Methodology

Faced with the challenge of investigating the relationship between a complex disease and multiple environmental factors within a 440-square-mile region, the Silent Spring Institute proposed developing a GIS to be used as the central data management and analysis tool for the study. We used the GIS to further define the problem of breast cancer on Cape Cod by conducting a detailed cancer surveillance effort. We also began to characterize the environment of the Cape, identify differences within the Cape, and explore relationships between the breast cancer incidence and environmental features.

In our cancer surveillance effort we used the GIS to geocode the addresses of 2,173 women with breast cancer for the period 1982 to 1994 (the full set of addresses currently available from the Cancer Registry). We also used residential land use data to refine estimates of populations for intercensal years. We used these refined population estimates to calculate standardized incidence ratios (SIRs) for towns, census tracts, and census block groups. The results of this cancer surveillance effort indicate that breast cancer incidence is about 20% higher on the Cape compared with the rest of Massachusetts.

When we looked at breast cancer incidence in geographic units smaller than the town, we identified six geographic areas scattered across the Cape where the excess cancer incidence is focused (Figure 1). We were particularly interested to note elevated incidence in the area of the Massachusetts Military Reservation (MMR), a Superfund site. A case-control study previously conducted in this part of the Cape by members of our research team also identified a statistically unstable association between breast cancer and the gun and mortar positions at the MMR (1). These sites were not only used for artillery practice, but also for burning propellant bags. Dinitrotoluene, a known mammary carcinogen in animals, is one of the chemicals found in the propellant. Other areas of elevated incidence are southern Falmouth; south Barnstable; a mid-Cape area including parts of Yarmouth, Dennis, and Harwich; a section of Orleans and Chatham; and south Truro.

We began using the GIS to explore how the environmental features in the areas of elevated incidence differ from the features in the rest of the Cape and in the rest of the state. We especially focused on environmental features that might be related to exposure to endocrine disrupting compounds (EDCs). EDCs include several chemicals that act like estrogens and are found in common commercial products and in the environment. Because breast cancer has been shown to be associated with lifetime exposure to natural estrogen, it is plausible that there might be a link between synthetic chemicals

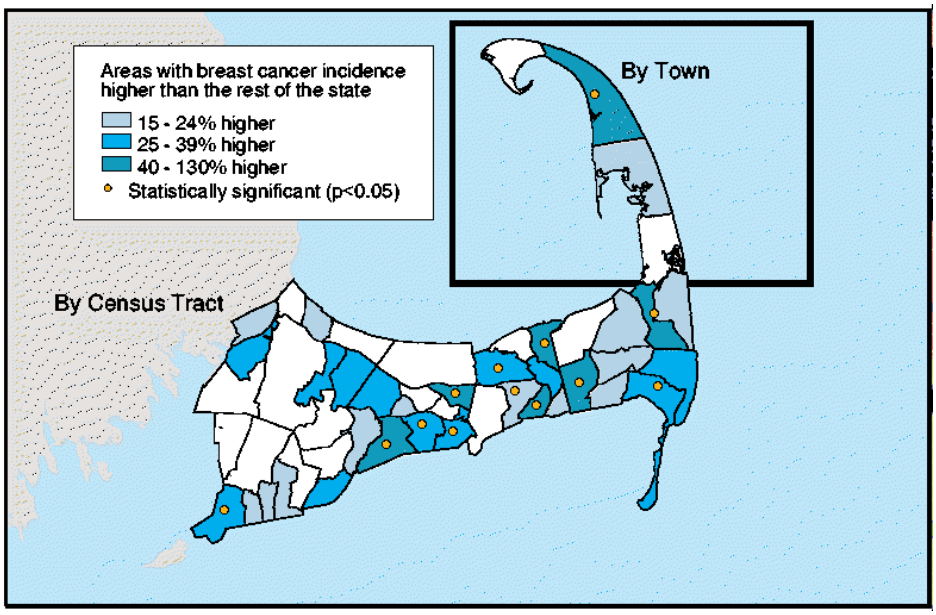

Figure 1 Breast cancer incidence by census tract, 1982-1994. Cape-wide breast cancer incidence was 20% higher than the rest of Massachusetts for this period. The circled areas are subregions of Cape Cod where the excess breast cancer incidence is concentrated.

that act like estrogen and breast cancer. We gathered data about two potential routes of exposure to EDCs on Cape Cod: exposure to pesticides through inhalation, dermal contact and ingestion, and exposure to drinking water impacted by waste water.

We used the GIS along with historical records and land use data to map areas of pesticide use. We focused on pesticides used on cranberry bogs and golf courses and those used to control tree pests. Our work using the GIS to study pesticides was described in a demonstration project for this conference.

Waste water has been shown to contain endocrine disrupting compounds. For several years, local, state, and federal agencies have been concerned about the impact of development on the Cape and of waste water disposal practices on the aquifer. The US Geological Survey collected data on analyses of private wells conducted by Barnstable County Health and Environment Department. Researchers have focused on nitrate-nitrogen as an indicator of water quality. Natural nitrate concentrations on the Cape are low, less than 2 mg/L. Waste water and fertilizer can cause nitrate concentrations to be elevated.

In an ecological epidemiology analysis we conducted using data from the cancer registry, we did not see any association between breast cancer incidence and elevated nitrate levels in drinking water. In the interest of getting the most out of readily available data, we explored the data further with statistical and data visualization software in order to generate hypotheses for further study.

We used the GIS to investigate how much the population and environmental features vary within the Cape and found substantial variation. One difference within the Cape that stands out is the distinction between the Lower Cape—the easternmost

part—and the rest of the Cape. The population density in this region, with the exception of Provincetown at the tip of Cape Cod, is lower than most of the rest of the Cape (Figure 2). The Lower Cape towns of Truro, Eastham, and Wellfleet are almost entirely dependent on private wells for their drinking water supply.

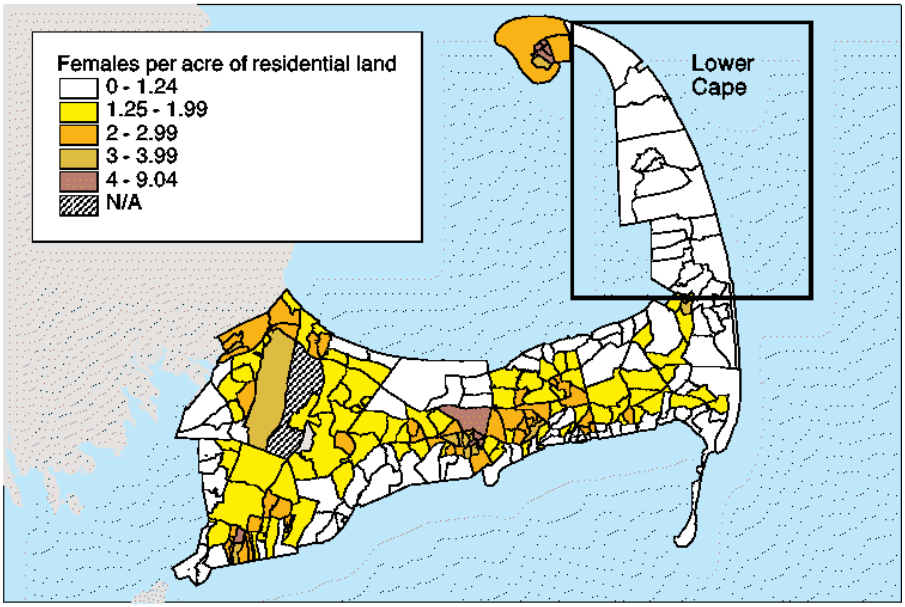

Figure 2 Population density by census block group, 1990. There are differences in both the environment and the population within the Cape. The towns of the Lower Cape, inside the rectangle, have lower population density.

Two factors that we speculated might be associated with water quality in areas served by private wells are housing density and housing age. Obviously, areas of dense housing would contribute more waste water to the aquifer. Older housing may include homes with cesspools and other waste water disposal systems that do not meet today's standards. In addition, the longer a septic tank is in operation, the greater the impact it will have on the aquifer. Housing density information was available from land use data and census data. Age of housing information was available from the census at the block group level.

Within the Lower Cape we found some variation in housing density (Figure 3) and age of housing (Figure 4). Denser and older residential areas were concentrated in the town centers. Wellfleet Center stands out as a location with particularly high density and old homes.

We created a series of scatter plots of SIRs by census block group versus percent housing greater than a specified number of years for those block groups of the Cape that are primarily dependent on private wells. We found that there did not appear to be an association when the x-axis was percent housing greater than 20 years. When we looked at percent housing greater than 50 years old there appeared to be a weak positive correlation (Figure 5). The correlation was strongest if only the census block groups

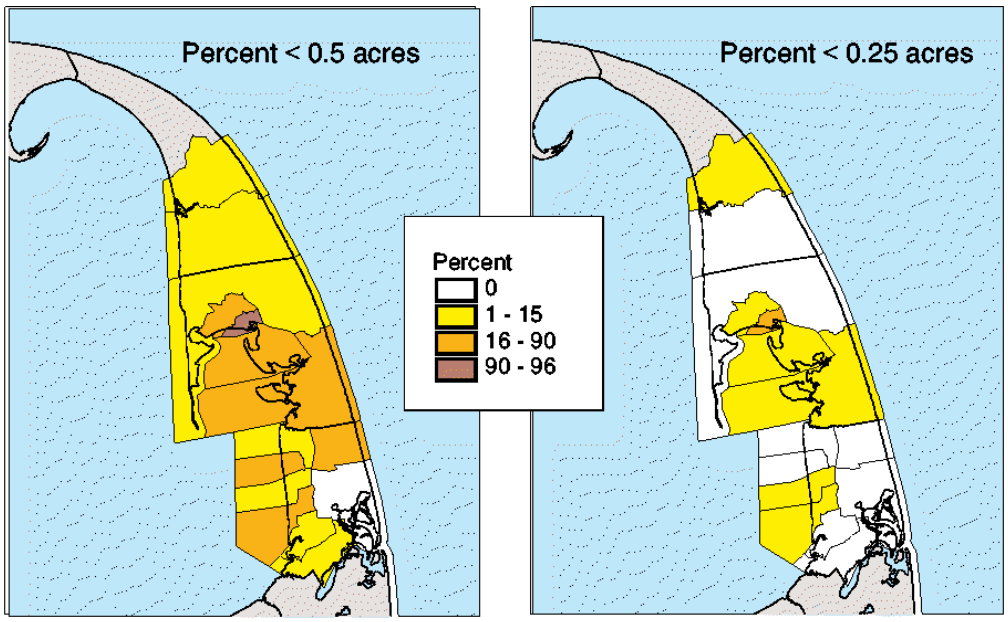

Figure 3 Percent high-density housing, 1971. Within the Lower Cape there are differences in housing density. Areas of higher density may have more wells impacted by waste water disposed of in septic tanks.

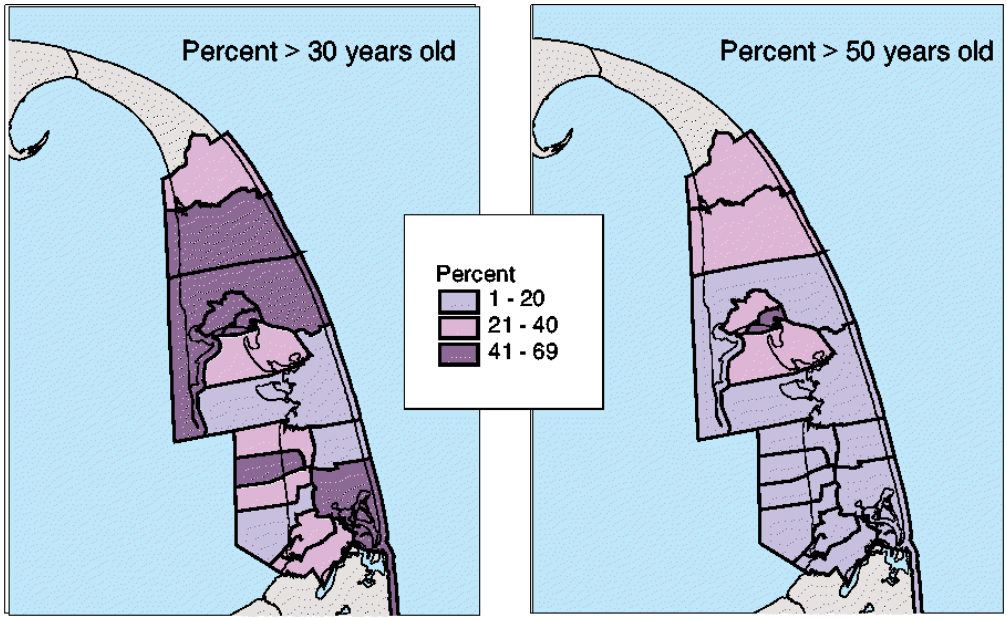

Figure 4 Percent old housing. The age of housing also varies within the Lower Cape. Wellfleet Center, in the middle of the figure, stands out as an area with old, dense housing.

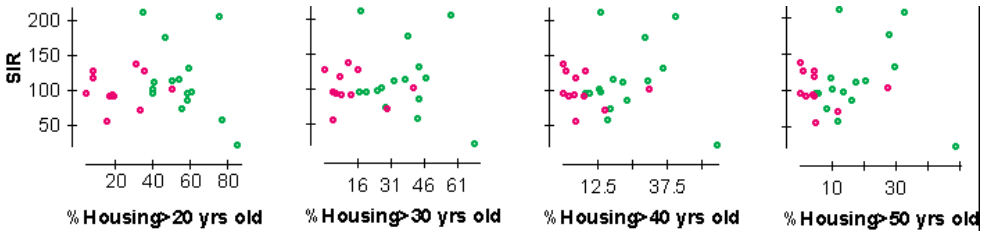

Figure 5 Breast cancer incidence and age of housing. Standardized incidence ratios (SIRs) by census block group were plotted against percent of old housing for the block groups where private wells are the primary source of drinking water. The correlation between breast cancer incidence and age of housing is highest when housing over 50 years old is used for the x-axis. The lighter dots are the block groups of the Lower Cape and the darker dots are the block groups for the rest of the Cape. Wellfleet Center stands out as a block group with low breast cancer incidence and old housing in the lower right of the plots.

of the Lower Cape were considered. Wellfleet Center stands out as an outlier with old homes and dense housing but low breast cancer incidence.

We also developed a simple scheme to take into account age of housing and housing density together. We calculated a housing risk index (HRI) according to the following formula:

$$HRI=a+d$$

where:
a=percent of old housing
d=percent of residential land in smaller than ½-acre lots up to a maximum value of 'a'

We assumed that if there was more dense housing than old housing then the excess dense housing must be new. We created a series of plots of SIRs versus the HRI (Figure 6). We found that the association between incidence and age of housing for the Lower Cape became even stronger when housing density was taken into account using the HRI.

Conclusion

We do not intend this sort of exploratory analysis to be used as evidence that certain environmental factors are responsible for the elevated breast cancer. A major limitation of this analysis is the lack of information available about confounding factors. For example, risk of breast cancer incidence has been shown to be correlated with elevated socioeconomic status (SES). It is possible that in the Lower Cape, the areas of older homes may be areas of higher SES because the older homes might be more desirable in this area. Ecological epidemiology analyses are intended to be hypothesis generating. In our case, the analyses suggest that the hypothesis that breast cancer incidence is associated with exposure to drinking water from shallow wells in areas of old and dense housing should be refined to focus on exposure in areas with high percentages of housing greater than 50 years old.

We are gathering data about confounding factors in a case-control study begun in the fall of 1998. GIS has been incorporated into the design of this study from the

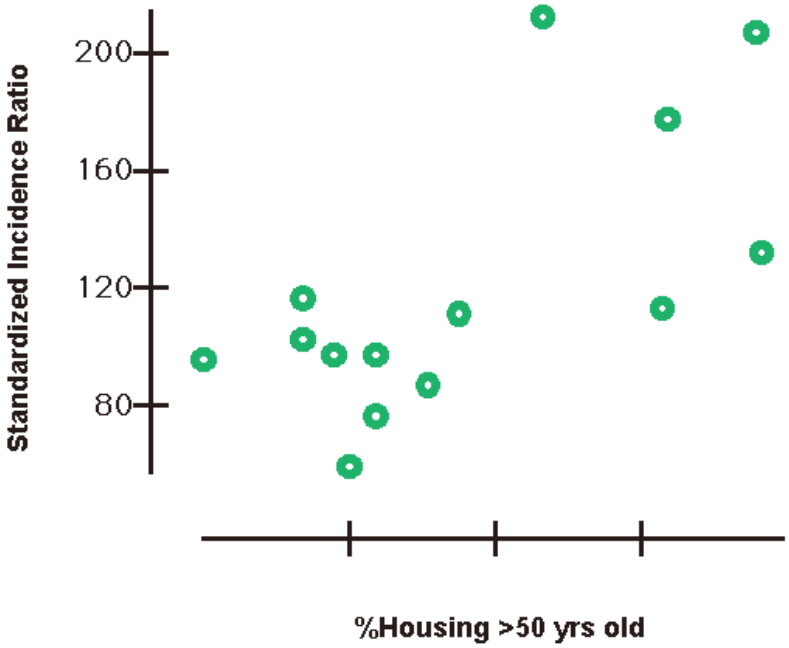

Figure 6 Breast cancer incidence, age of housing and housing density in the Lower Cape. The correlation between breast cancer incidence and age of housing is stronger when the age of housing is weighted to account for housing density. Wellfleet Center was excluded from this plot.

beginning. GIS data compiled in the first phase of our study are being used to develop exposure variables in the current case-control phase. The exploratory analyses we conducted in the first phase of our study illustrate some of the kinds of analyses that can be done when GIS data are combined with disease incidence data.

Acknowledgements

This work was supported by funds appropriated by the Massachusetts Legislature and administered by the Massachusetts Department of Public Health under Contract DPH79007214H11, and by the generous contributions of the Susan G. Komen Breast Cancer Foundation/Boston Race for the Cure.

Reference

1. Aschengrau A, Ozonoff DM. 1991. *The Upper Cape cancer incidence study: Final report*. Boston, MA: Boston University School of Public Health. September.